An NK Landscape Based Model Mimicking the Protein Inverse Folding Problem

Sune Nielsen[1], Gregoire Danoy[1], Pascal Bouvry[1], and El-Ghazali Talbi[2]

[1]FSTC/CSC, University of Luxembourg, Luxembourg
[2]University of Lille - INRIA - CNRS, France

1 Abstract

This work introduces a new NK Landscape based model instance designed to mimic the properties of one challenging problem in biology: the Inverse Folding Problem (IFP). Protein structure prediction is an essential step in understanding the molecular mechanisms of living cells with widespread application in biotechnology and health. Given a protein as reference and its corresponding tertiary (3D) structure, the IFP consists in finding RNA sequences which produce very similar 3D structures. In this work an IFP model previously developed to match a neural network secondary structure prediction with a reference, is used. Numerous landscape analysis tests of a problem instance based on the protein 1b3a are conducted. The results are then used to parameterise the new model - a combination of two NK Models, with different K and neighborhood definitions. The model can accumulate the characteristics of both its underlying NK Models to match the epistatic interactions and landscape walk features of the IFP. This work is an initial step in the creation of a fast benchmark for all algorithms targeting protein sequence optimisation. With a simple definition based on the well-known NKModel, the motivation is to make the IFP problem more accessible to optimisation specialists and model experts.Furthermore the statistical nature of the NK Model may provide the ground for a theoretical estimate on the number of protein sequences which fold into a given protein structure.